8. Communicating Science to the Public: A Comparison of Lexicogrammatical Features in Student-Produced and Popular Science Writing

Jordan Batchelor
INDEPENDENT RESEARCHER

Jordan Smith
UNIVERSITY OF NORTHERN COLORADO

Abstract: Technical communicators often communicate scientific findings to a broad audience, but little empirical research has analyzed the linguistic features of popularized science texts. We compare the lexicogrammatical features of student-written scientific reports with science press releases (SPRs). We compared a corpus of SPRs with a corpus of student-written scientific reports using key feature analysis (Biber & Egbert, 2018), a statistical method that measures the effect sizes of different lexicogrammatical features across the corpora. The features identified as key for both corpora were grouped according to their functional themes. Science press releases used lexicogrammatical features related to information density, informational reports of past events, and timeliness. Student-written science reports used lexicogrammatical features related to expression of stance, scientific style, coherence and development of arguments, and writing prompt. Based on our findings, we provide six recommendations for teachers, students, and practitioners who want to mimic the style of SPRs as a form of science popularization: (1) tell stories with active voice, (2) include others' words by using quotations and communication verbs, (3) avoid technical jargon, (4) use adjectives of time to highlight the importance of research, (5) minimize expressions of authorial stance, and (6) report facts rather than make arguments.

Keywords: Science communication, popularization, press releases, corpus linguistics, key feature analysis

One of the important tasks that technical and professional communicators perform is communicating scientific findings to a general audience (Carradini, 2020; Mogull, 2018). While many scholars in the field have studied public communication of science in different modes, including online forums (Cagle & Herndl, 2019), film (Spoel et al., 2008), and press releases (Weber, 2020), little empirical

research has been done to better understand linguistic variation in popular vs. academic science writing. Such research is needed because it can be used to inform best practices for communicating science to the public.

In this chapter, we address this need for more empirical work by analyzing the lexicogrammatical features associated with two registers of scientific writing: student-written reports and scientific press releases. *Lexicogrammar* is a term from functional linguistics that "[emphasizes] the interdependence of vocabulary (lexis) and syntax (grammar)" (Nordquist, 2020, para. 1). The results of the research presented in this article contribute to a description of the ways in which popular science writing (in the form of press releases) differs from the more traditional scientific reports students often produce. These results can help instructors, scientists, and science communicators better understand how to write in ways that meet existing expectations for writing toward nonexpert audiences.

■ Literature Review

In this section, we review the need for students and practitioners to communicate to popular audiences. Traditionally, science students are taught to produce a limited number of scientific text types such as methodology recounts and research reports; however, these texts are often not accessible to nonexperts. Teaching these students—many of whom are required to take technical communication classes—how to popularize their work can help them learn to communicate scientific information that more people can understand. Teaching popular-science-writing principles can also help aspiring technical communicators to more effectively communicate information from subject-matter experts. We focus in this study on the press release as an example of popularized science writing. We further elaborate each of these points in turn below.

■ The Need to Communicate Science to Nonexpert Audiences

Science and technology are crucial aspects of daily life. In the US, the number of college students majoring in science, technology, engineering, and mathematics (STEM) is increasing, in part because STEM careers represent a significant proportion of the U.S. labor force (Burke, Okrent, & Hale, 2022). Academic research is also growing rapidly. Today, more research articles are being published in academic journals than ever before (Hyland & Jiang, 2019). Moreover, recent issues such as the COVID-19 pandemic, the growing threat of climate change, and advancements in artificial intelligence have brought science and its communication to the public to the fore. In short, the 21st century is one marked by the production of scientific information and the need to communicate it.

However, the language that scientists use to communicate with one another, in particular through research writing, does not cater to nonexperts (Bazerman, 1988; Halliday & Martin, 1993). Because of this, when communicated with broad,

diverse audiences, scientific information often undergoes a process of "popularization." Put simply, science popularization refers to the process of communicating scientific information to nonexperts, usually resulting in some local and global revisions to the content (Gotti, 2014). Although the stereotypical view of popular science sees its audience as being passive and incapable of understanding authentic science (Hilgartner, 1990), contemporary scholars view the process as more complicated (Myers, 2003; Paul, 2004). Popular science is not a simplification of real science but a "discursive reconstruction" (De Oliveira & Pagano, 2006, p. 628) and "recontextualization" (Calsamiglia & Van Dijk, 2004, p. 370) of scientific information for a new rhetorical context. Thus, popularizing science requires not just revising its linguistic features, such as jargon and syntax, but also revising its rhetorical features as well, such as purpose and organization (Fahnestock, 1986; Hyland, 2010).

Three decades ago, Ann Mauzy (1994) described the need for technical communicators to communicate complex information to diverse, nonexpert audiences: "The stakes…are high," she wrote. "Communicators need to target the public as the *most important audience* and to urge scientists to do the same" (p. 143, emphasis added). This need to prioritize the public is only increasing. Web 2.0 technologies have broadened the methods by which scientists and technical communicators can reach lay audiences (Luzón & Pérez-Llantada, 2019). Social media, science blogs, TED Talks, and science podcasts are examples of these expanding contexts. Traditional methods used to popularize science, too, remain crucial. Sharon Dunwoody (2014) argued that science journalism represents an important source of "independent, evidence-based information" in an age where ordinary citizens are bombarded with one issue after another (p. 27). Indeed, global scientists regularly communicate with journalists to share their research (Peters et al., 2008), and the internet allows for faster and cheaper publication of science news online.

However, more research is needed to better understand the skills involved in popularizing science. Particularly if we are to train aspiring technical communicators—as well as the many STEM students who regularly enroll in technical communication courses—to engage in this process, practical insights into the linguistic features of popular-science discourse would aid in developing materials for training students how to share scientific research with broader audiences. Below, we home in on one form of popular science, namely press releases sharing science news, and the kinds of writing that college-level students are expected to produce. Ultimately, we argue that examining the linguistic and functional similarities and differences between these two kinds of discourse can provide helpful insight for researchers, instructors, and students of technical communication.

Science Genres and Student Writing

Professional scientific activities demand vehicles through which their processes and results can be effectively communicated to others. One way of describing

these vehicles is through genre, or the "ways of recognizing, responding to, acting meaningfully and consequentially within, and helping to reproduce recurrent situations" (Bawarshi & Reiff, 2010, p. 3). Naturally, there is a plethora of genres frequently used to communicate science, but the most influential and prestigious has been the research article (RA) (Hyland & Salager-Meyer, 2008; Samraj, 2016). Charles Bazerman (1988) examined the earliest developments of the RA (called "experimental reports"), illustrating the connection between generic form and social function. For example, as reports took on an argumentative function, serving to adjudicate disputes and generalize knowledge, writers began to situate their work within a longer scientific conversation more fully, include more precise methodological description, and write longer articles. John Swales's influential work on move analysis (e.g., Swales, 1981, 1990), a framework which seeks to uncover the communicative functions motivating portions of text, also bolstered the study of the RA and its role as the premiere site of the investigation of scientific discourse.

From a linguistic perspective, M. A. K. Halliday and J. R. Martin's (1993) influential work on scientific discourse highlights the science writer's tendency to turn parts-of-speech like verbs and adjectives into nouns (called "grammatical metaphor"), a process which developed over time to suit the needs of communicating empirical work to technical audiences. Others too have noted the nominal style of scientific writing (Wells, 1960), its tendency to create lengthy, complex noun phrases by way of stacked modifiers (Biber et al., 1999, Ch. 8), and the "clearer" style bestowed by maintaining actions as verbs (Williams, 2009). Moreover, these features are not just stylistic but also functional. Douglas Biber and Bethany Gray (2016) argued that this style allows for compact, informationally dense prose useful when writing for increasingly specialized audiences (p. 207). In short, scientific discourse functions to effectively communicate science to other professionals, and the RA is the premiere genre used to communicate it.

Students studying in colleges and universities must also contend with the RA. Susanne Pelger and Pernilla Nilsson (2016, p. 440) stated that students in science classes are often expected to write empirical reports in the style of a scientific paper, a fact which they see as problematic, since it limits the students' skills in communicating science, especially to diverse audiences. This makes some sense. A key characteristic of empirical reports is their technical audience (Gotti, 2014). And yet writing toward general audiences has pedagogical benefits, as well. Being able to recognize the similarities and differences of popular and professional science discourse reflects a learner's awareness of their rhetorical and linguistic characteristics (Hyland, 2010; Parkinson & Adendorff, 2004). The STEM students in Wendy Crone et al. (2011), Pelger (2018), and Pelger & Nilsson (2016) were found to improve audience awareness and develop a new perspective on their own research as a result of popularizing their own research.

We argue that this benefit can extend beyond science students to technical writers as well, and that the science press release is one text type through which

students can gain such practice. Next, we discuss the theoretical underpinnings of the study, namely register analysis, and how it can be applied to a study of the language of press releases and student writing.

From Genre to Register

The current study seeks to examine features of the science press release and compare it to student-produced science writing. The features we examine reflect the theoretical lens we adopt. Specifically, we adopt a register-studies perspective by examining lexicogrammatical features of two corpora of texts.

Put plainly, register analysis emphasizes the lexical and grammatical features of texts and how they vary across text types. While genre studies often examine longer stretches of text like rhetorical moves (e.g., Swales, 1990), register studies focus on the variation of linguistic features between text types and attempt to connect those features to the situational context in which the texts were produced (Biber & Conrad, 2019). In other words, linguistic choices are not just formal but also functional, reflecting the situational characteristics of discourse like audience, purpose, topic, modality, and more. We see this approach as complementary to the already substantial body of genre work on science and non-science written genres.

We compare student-produced science writing with science press releases in order to understand which lexicogrammatical features distinguish the press releases from the scientific reports that many students are taught to write. Even though these text types differ with respect to the audiences they aim to reach, they both share the same primary communicative purpose, namely to disseminate scientific information. For this reason, we argue that the comparison undertaken in this study is valid and capable of producing useful results. The results of this analysis allow us to explicitly teach these linguistic features to students when helping them learn how to write science press releases as a way to communicate science to a broader audience.

In the next section, we review relevant background literature on the press release, before presenting our research questions.

Communicating Science through Press Releases

Scott Mogull (2018) argued that technical communication teachers need to teach students to write for broad audiences and even advocated for technical communication classrooms to act as places for scientists to learn how to popularize their work (2011). One text type that teachers can use to help their students learn how to write science popularizations is the press release. While a selection from the ever-expanding catalog of new-media registers might seem more relevant or exciting to teach students, traditional sources of news media such as the press release continue to play an important role in the process

of communicating science. According to the National Academies of Sciences, Engineering, and Medicine (2017),

> Despite the growing impact of new media, much of the scientific information Americans receive through media still originates from traditional journalism, including information transmitted via links on social media. Therefore, science communicators need to understand the tools used by journalists to shape scientific information, especially that related to contentious societal issues. (p. 69)

In addition, technical communication scholars have identified the press release as a text type that students should learn to write. For example, Jon Balzotti and Derek Hansen (2019) asked their students to write a press release as part of a pedagogical case study, and Mogull (2018) devoted a whole chapter on writing press releases in his guide on science and medical writing.

Press releases exist between two different fields—media studies and professional communication (Jacobs, 2014)—and have two primary purposes: to promote and to inform. Paola Catenaccio (2008) described this dual purpose, noting that while press releases must contain informative material, they must also simultaneously "persuade journalists that they are newsworthy [and] they must persuade the general public that the company is profitable/trustworthy/ offers something they need, etc." (p. 14). The reach of press releases is broad because they are taken up by journalists who report on the findings described in the press release (Granado, 2011; Maat, 2007; Jacobs, 1999) and sometimes publish much of the language of the press release itself. Research has shown that scientific medical articles that are accompanied by a press release are more likely to receive news coverage (Stryker, 2002), and high-quality press releases result in higher-quality coverage (Schwartz et al., 2012). This second finding offers further evidence that journalists do in fact rely on press releases when writing about scientific research.

Studies have discussed the importance of press releases and the role they play in communicating scientific findings to a wide audience. For example, research on the content of scientific press releases has found evidence for exaggerated claims (Sumner et al., 2016) and helping the public gain a useful albeit incomplete glimpse into science in the making (Weber, 2020). Other studies have investigated pragmatic aspects of press releases including the use of metadiscourse markers (Liu & Zhang, 2021) and metapragmatic features like self-reference and pseudo-quotation (Jacobs, 2014), while still others have examined the relationship of the press release to other, related texts in its "genre chain" (e.g., Maat, 2007). However, few studies have analyzed press releases on a lexicogrammatical level. One exception is Pititul Chavanachid and Passapong Sripicharn (2019) who analyzed variation in linguistic features and move structures of press releases. We argue that register analysis provides a sound framework for analyzing the lexicogrammatical features characteristic of science press releases.

■ A Register Analysis of Science Press Releases and Student Writing

We adopt a register perspective to survey the lexicogrammatical features descriptive of science press releases and student writing as they relate to one another. The aim is to provide empirical insight into how science press releases are written to appeal to broader audiences, insights which may prove useful for both practitioners and researchers.

With these details in mind, the current study seeks to answer the following research questions:

> **RQ1:** What lexicogrammatical features are descriptive of science press releases relative to student writing?
>
> **RQ2:** What lexicogrammatical features are descriptive of student writing relative to science press releases?

Although these questions are largely linguistic, we also seek to describe how their answers can inform our understanding of the communicative functions of the registers. Linguistic analysis coupled with a description of the available situational characteristics of the texts being analyzed is one method for accomplishing this goal (Biber & Conrad, 2019).

■ Methods

To answer RQ1 and RQ2, we used a statistical method from corpus linguistics called key feature analysis (Biber & Egbert, 2018), which compares frequencies of about 150 lexicogrammatical features in texts from the two registers. To better understand how those features relate to communicative functions, we grouped the features identified as key for each register and interpreted them functionally. In this section, we describe how we collected our corpora, and we provide a detailed description of key feature analysis. In the following section, we present the thematic groups of key features and their functional interpretations.

■ Corpus Collection

To answer this study's research questions, we collected two corpora. The first corpus consists of press releases reporting on published scientific research. We call this corpus the science press release (SPR) corpus. The second corpus consists of undergraduate and graduate student science writing. We call this corpus the MICUSP subcorpus, since it consists of a subset of the student writing publicly available from the Michigan Corpus of Upper-Level Student Papers (MICUSP) (The Regents of the University of Michigan, 2009).

The SPR corpus includes 2,943 press releases produced by institutions that report on published academic research, mainly U.S. universities, scientific societies,

and health organizations. All texts were collected from the "news releases" section of EurekAlert!, a nonprofit news organization operated by the American Association for the Advancement of Science (AAAS). Text collection was performed in December 2022, and all texts were published in the year 2021 or 2022. An initial, larger corpus was collected and then cleaned by removing picture/video articles, duplicated texts, and texts less than 300 words or longer than 2,000 words. The mean text length in the corpus is 700 words (SD = 252).

The MICUSP subcorpus includes 153 texts collected from the online repository of MICUSP, a corpus collected in 2009 to represent A-level papers written by University of Michigan students across discipline, genre, and level of education (The Regents of the University of Michigan, 2009). These student papers stem from a variety of classrooms and writing prompts, but all texts included in our sub-corpus were written by final-year undergraduate and first-, second-, and third-year graduate students from one of seven disciplines, namely natural resources and environment, biology, nursing, civil and environmental engineering, industrial and operations engineering, physics, and mechanical engineering. All texts were classified as "reports" by the corpus collectors and include textual features like headings and subheadings, citations and references, figures and tables, and formulae. In general, these student papers are informational, with the main focus being to objectively report on a chosen scientific topic. (An exemplar text can be found at https://elicorpora.info/view?pid=BIO.G0.02.1.) The mean text length in the corpus is 2,667 words (SD = 1,788). Table 8.1 summarizes the descriptive information of the two corpora.

As Table 8.1 shows, the SPR corpus has about 2 million words across about 3,000 texts. The MICUSP subcorpus has fewer but longer texts, resulting in about 400,000 words across 153 texts.

Table 8.1. Descriptive Information for the Scientific Press Release (SPR) Corpus and MICUSP*

	SPR Corpus	MICUSP Subcorpus	Total
# of texts	2,943	153	3,096
Years of publication	2021–2022	2009	2009, 2021–2022
# of words	2,061,562	408,019	2,469,581
Mean text length	700	2,667	–
SD text length	252	1,788	–

*Descriptive information about the scientific press release (SPR) corpus and the MICUSP subcorpus used in this study. The texts in both corpora focus on scientific topics, though they are written for different purposes, which accounts to some to degree for the variation in the corpora: the SPR corpus is composed of a higher number of shorter texts, while the MICUSP subcorpus is composed of a lower number of longer texts.

Texts in the MICUSP subcorpus were written a little over ten years prior to the press releases. The texts in both corpora generally discuss scientific topics, though their more specific purposes differ. SPRs are written for wide audiences and function to report on recently published research from researchers within the writer's organization. The MICUSP student papers are written for an instructor and can report on either empirical or nonempirical work. Some MICUSP papers have subsections like introductions, methods, and findings, while SPRs generally lack such distinctions. In short, we expect that differences in form and content between these corpora will result in the reliance on different linguistic features, which we describe next.

Key Feature Analysis

We adopted key feature analysis (Biber & Egbert, 2018; Egbert & Biber, 2020) to investigate the lexicogrammatical profiles of the two corpora. Key feature analysis involves using computer programs to part-of-speech tag and syntactically parse the words and clauses of texts. We tagged our corpora using the Biber Tagger (Biber, 1988), which tags texts for about 150 different grammatical and lexicogrammatical features. To give an example of the features that the Biber Tagger identifies, five features, their tags, and examples are shown in Table 8.2.

Table 8.2. Five Features Identified by the Biber Tagger*

Feature Tag	Full Name	Examples
mod_poss	Modal of possibility	*can, may, might, could*
vb_past	Past tense verb	*said, was, ate*
vb_mental	Mental verb	*think, know, believe*
conj_advl	Adverbial conjuncts	*however, therefore, thus*
passive_short	Short passive voice verb	any finite passive verb phrase without a stated agent

A sample of five grammatical and lexicogrammatical features that can be identified using the Biber Tagger. These features include simple grammatical features (e.g., past-tense verbs) as well as functional lexicogrammatical features (e.g., short passives, like The test was taken.*)

The Biber Tagger provides normalized frequencies of all features in each text. Normalizing raw frequencies is important because it allows the researcher to compare the occurrence of features across texts and corpora of different lengths. In this study, normalization was calculated as follows:

$$normalized\ frequency = \frac{Raw\ frequency}{Length\ of\ text} \times 1000$$

For example, if 7 modal verbs occurred in a press release of 600 words, the normalized frequency in that text would be 11.67 per 1,000 words (ptw). Normalized frequencies are calculated per text rather than per corpus to allow the researcher to apply inferential statistical tests, which often require means and standard deviations.

Next, we describe the statistical analysis applied to the tagged corpora.

▉ Statistical Analysis

Key feature analysis involves comparing the standardized effect sizes of certain features across corpora. The standardized effect size adopted here is Cohen's *d* (Cohen, 1988). Although many effect sizes exist (e.g., Hedge's *g*), we chose Cohen's *d* because our sample sizes were large enough to avoid bias (see Lin & Aloe, 2021), and because of its ubiquity in key feature analysis (e.g., Biber & Egbert, 2018) and corpus linguistics (e.g., Brezina, 2018). Cohen's *d* is defined as follows:

$$d = \frac{M1 - M2}{\sqrt{\frac{SD1^2 + SD2^2}{2}}}$$

where, "M" refers to the mean normalized frequency of a feature, "SD" refers to its standard deviation, and "1" and "2" refer to the reference corpus and comparison corpus, respectively. In written English, *d* is calculated by subtracting the mean of the reference corpus from the mean of the comparison corpus, which is then divided by the square root of the sum of the squared standard deviations divided by two. As an example, if the frequency of modal verbs in Corpus #1 is *M* = 11.67 and *SD* = 5, and the frequency of modal verbs in Corpus #2 is *M* = 7.5 and *SD* = 3.5, then *d* = 0.97.

Cohen's *d* can be positive or negative and is usually interpreted as follows: a large effect is ±0.8, a medium effect is ±0.5, and a small effect is ±0.2. Biber and Jesse Egbert (2018) examined all key features with *d* > ±0.3, while Egbert and Biber (2020) examined key features with *d* > ±0.8. Because the SPR corpus and MICUSP subcorpus share more similarities than differences, we did not expect to encounter particularly large *d* values. Thus, we set the cut-off point at *d* > ±0.5, or a medium effect, which Jacob Cohen (1988) considered to be "one large enough to be visible to the naked eye" (p. 26). A qualitative review of our key features with medium and large *d* values also suggested that this cut-off value captured important differences between the corpora.

Cohen's *d* values were calculated using the *lsr* package (Navarro, 2015) in R (R Core Team, 2020). Data visuals were produced using the *ggplot2* package (Wickham, 2016).

▉ Results and Discussion

In this section, we report on the findings of the key feature analysis. First, we provide an overview of the key features. Then we explore the positive key features of the SPR

corpus and the MICUSP subcorpus. Excerpts are provided to illustrate the key differences between the language of science press releases and student science writing.

■ Key Features of the SPR Corpus and MICUSP Subcorpus

Twenty-two key features with Cohen's *d* values larger than 0.5 or smaller than -0.5 were identified between the SPR corpus and MICUSP subcorpus. This means that these features are most key when examining how the two corpora differ with respect to their linguistic profiles. Nine of these features had positive Cohen's *d* values, meaning that they were more frequent in the SPR corpus. The remaining 13 had negative Cohen's *d* values, indicating that they were more frequent in the MICUSP subcorpus. Figure 8.1 visually compares these features.

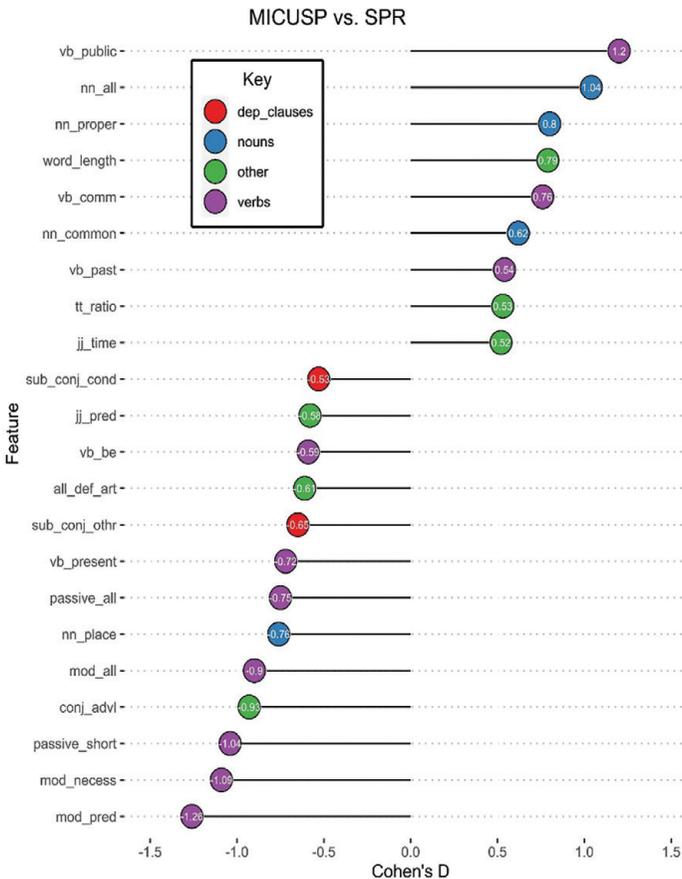*Figure 8.1. A visual comparison of the key features in the SPR corpus and the MICUSP subcorpus. The nine features with positive Cohen's d values are shown with lines extending to the right. The 13 features with negative Cohen's d values are shown with lines extending to the left.*

Figure 8.1 shows that features referring to verbs, such as prediction modals (mod_pred), short passives (passive_short), and present tense verbs (vb_present), represent the most frequent structural category. Most verbal features were descriptive of the student writing, though public verbs (vb_public) were notably frequent in the press releases. Four key features referred to nominal structures, three of which were descriptive of the press releases, including common and proper nouns. Two types of dependent clauses were more frequent in the student writing, including conditional dependent clauses (sub_conj_cond; e.g., *if…then…*) and dependent clauses expressing other semantic categories, such as time, place, and manner (sub_conj_othr; see Biber et al., 1999, pp. 818–819 for examples). A range of other structures were also found to be key features, including adjectives of time (jj_time) and definite articles (all_def_art) in the press releases, and predicative adjectives (jj_pred) and linking adverbials (conj_advl) in the student writing. Finally, word length (word_length) and type-token ratio (tt_ratio) were key features of the press releases. These features are measured on a continuous scale using decimals, so a larger Cohen's *d* value means longer words and greater lexical diversity.

Below, we examine these key features more closely by looking at their uses in discourse. First, we explore the positive key features descriptive of science press releases before examining the key features of the student science writing.

◼ Positive Key Features of Science Press Releases

Examining the key features with larger, positive Cohen's *d* values helps to highlight the important lexicogrammatical features frequent in SPRs. Table 8.3 organizes the nine key features of the science press releases thematically. The labels for each group represent our functional interpretations of the features.

Table 8.3 presents three thematic groups: information density, informational reports of past events, and timeliness. We discuss each theme in turn below.

Table 8.3. Key Features of the SPR Corpus Relative to the MICUSP Subcorpus, Grouped Thematically*

Theme	Features
Information density	Type token ratio, word length
Informational reports of past events	Public verbs and verbs of communication, past tense verbs, all nouns, proper nouns, common nouns
Timeliness	Adjectives of time

The key features of the SPR corpus when compared against the MICUSP subcorpus. The features (shown in the right column) were grouped thematically and given a label (shown the left column).

Information Density

Information density includes two features, namely type-token ratio (TTR) and word length. TTR refers to the ratio of unique words in a text to all words in a text. The greater the number of different words in a text, the larger the TTR and more lexically diverse that text is. Similarly, word length speaks to the lexical characteristics of texts by measuring the average number of letters in a given word. In this study, the science press releases tended to be more lexically diverse and have longer words; thus, TTR and word length were positive key features. The following excerpt from an SPR exemplifies these features.

> **Text Sample #1:** Science Press Release
>
> If the crosslinking occurs in water, the material can then simply be dried in a mold. The result is a tough, transparent, solvent-resistant bioplastic. Its mechanical properties can be varied by changing the proportion of PEG. This allows for the production of bioplastics with high mechanical strength at room temperature in any shape desired, and without toxic chemicals or complex processing steps such as liquefaction, extrusion, or blow molding. Their breaking stress exceeds those of many commercial plastics. One problem left is that they swell in water. If ELP is crosslinked in a water/glycerol solution, the material gels into soft, elastic bioplastics.

The press release from which Text Sample #1 comes has an overall TTR of 38, meaning that 38% of its words were unique (i.e., not repeated), and an average word length of 5.6 letters. The above excerpt, taken from the longer text, has a TTR of 77 and an average word length of 5.9 letters. As the excerpt suggests, SPRs with higher TTR and long words often perform the function of explanation. That is, the writer explains a scientific concept, the procedures of a research study, or some other detail, resulting in a more diverse (and sometimes technical) vocabulary with more letters. This purpose contrasts with the reporting of others' words and ideas, which is the main function of the second thematic group of key features.

At the same time, readers are unlikely to be significantly affected by the longer words and greater lexical diversity of the SPRs. The mean TTR of the SPR corpus is 32.2 (SD = 2.16) and the mean TTR for the MICUSP subcorpus is 31 (SD = 2.42). Because the SD values are small relative to the mean values, which are not identical in size, the resulting Cohen's d is large enough to be a key feature. A similar point can be made for word length, which represents a medium effect size despite showing practically small differences (M = 5.41 letters per word compared to M = 5.19). Thus, while these differences may have a cumulative effect when viewing texts from the perspective of a corpus, they are nonetheless practically small, representing mere percentage point differences between the two text types.

Informational Reports of Past Events

The four key features listed with "informational reports of past events" in Table 8.3 relate more clearly to the functions of science press releases. The title given to this thematic group is borrowed from Biber, Mark Davies, James Jones and Nicole Tracy-Ventura's (2006) study of written and spoken Spanish-language texts. The authors identified a cluster of linguistic features, including proper nouns, past tense verbs, long words, and attributive adjectives, in a corpus of literate text types like business letters and newspaper reportage. The authors showed that these features were often used to report past events, which can also be seen in this study. Proper nouns and past tense verbs were key features of the SPRs, while adjectives of time and longer words were also present but listed under other thematic headings. To the features functioning to report past events, we also add common nouns and verbs of communication.

Informational reports of past events in press releases are characterized by direct and indirect reports of others' words and ideas, utilizing especially past tense verbs documenting study procedures and verbs of communication. Sentences with these features often have human subjects and descriptions of their academic and professional affiliations, resulting in more proper nouns. The following excerpt from a science press release illustrates these features. Verbs of communication are bolded, past tense verbs are italicized, and proper nouns are underlined.

> **Text Sample #2:** Science Press Release
>
> In obese mice whose tumor stress granules *were* blocked, 40% *were* cancer free after 300 days, with no sign of the cancer anywhere in the animals' bodies. "This magnitude of response is extraordinarily rare," **says** <u>Dr. Grabocka</u>. These experiments *showed* that stress granules *were*n't just present in cancer cells, they *were* actually driving the growth of cancer at the very start. "This is the first direct evidence linking stress granules to cancer progression," **says** <u>Dr. Grabocka</u>. Importantly, <u>Dr. Grabocka's</u> lab also *identified* drug targets that could stop the formation of stress granules in obesity-related pancreatic cancer. The next steps are to test existing small-molecule inhibitors to see if they can be translated for use in humans. "Conditions of cellular stress, like obesity, increase the number of stress granules present in cells, and may drive the formation of pancreatic and other cancers," **says** <u>Dr. Grabocka</u>.

Text Sample #2 represents a typical example of report writing in an SPR. The writer weaves direct quotations from the study's author, often using the verb *say*, with details of the study's procedures and findings, often in the past tense. While not shown in this short excerpt, human actors are first introduced along with their affiliations, after which their names are repeated along with other proper nouns referring to academic journals (e.g., *Nature, Cell*), locations (e.g., *US, China*), and certain technical vocabulary (e.g., *Alzheimer's, Parkinson's*).

Among this discourse is a variety of common nouns. Some of the most frequent are nouns referring to general research processes, including inanimate (e.g., *research, science, data, result, finding*) and animate (e.g., *author, team, professor, group, scientist*) words. Some common nouns are slightly more technical, referring especially to the health and biological sciences (e.g., *cell, cancer, disease, medicine, brain*). The presence of these common nouns suggest two characteristics of SPRs. First, the greater presence of nouns is often attributed to informationally dense, highly literate registers (Biber & Gray, 2016). At the same time, the key nominal features in the SPR corpus are decidedly nontechnical, referring to named entities and common nouns. As a result, SPRs represent an informationally dense register that is also accessible to many, as it narrates the scientific process by focusing on people, their past actions, and broad details of the research.

▪ Timeliness

The final key feature discussed here is adjectives of time, which are adjectives that refer to chronology, age, and frequency (e.g., *new, young, rare*). It might seem that many of these adjectives would also be useful for reporting past events, as they add contextual detail to the statements of press releases. For example, adjectives like *older, old, past, prior*, and *late* were used by SPR writers to mark events or situations as occurring in the past. However, the primary reason for grouping adjectives of time under the theme of "timeliness" is because the most frequent adjective in the SPR corpus was *new*, which was more than twice as frequent as the second most frequent adjective, *national*. Indeed, over 70% of the SPRs (N = 2,098) included at least one instance of the adjective *new*, often near the beginning of the article. The high frequency of *new*, along with certain other adjectives (e.g., *current, recent, present, increasing, latest*), suggests that an important function of science press releases is to make new research appear timely and relevant. The following excerpt illustrates this use of *new* (bolded), while other words contributing to the sense of timeliness are also underlined.

> **Text Sample #3:** Science Press Release
>
> As people transitioned to working from home at the onset of the coronavirus pandemic, journal submissions from academics increased across the board. But a **new** study from Northwestern University found as men's scholarly productivity increased, women physicians were submitting less. The research reflects wider trends in academic publishing and is the first study to find such patterns in family medicine. The study contributes to a growing body of evidence that the pandemic caused unique career disruptions for women as they became stretched thin during remote work, causing stress, burnout and anxiety.

This excerpt represents the first four sentences of its respective press release. The first sentence contextualizes the not-yet-introduced study by mentioning the effects

of the COVID-19 pandemic on working trends and then transitioning into a sentence introducing the "new study," which fits within "wider trends" in academia. The study is then referred to as the "first" one to accomplish what it did. In short, certain adjectives of time often coincide with other features, which together form a discourse of timeliness that helps to sell a study, its authors, and its findings to readers.

Positive Key Features of Student Science Writing

The previous section focused on the nine key features descriptive of the SPR corpus. In this section, we turn to the 13 key features descriptive of the MICUSP student papers. Table 8.4 organizes these 13 features into four thematic groups with labels representing the rhetorical functions of each one.

Table 8.4. Key Features of the MICUSP Sub-Corpus Relative
to the SPR Corpus, Grouped Thematically*

Theme	Features
Expression of stance	Prediction modals, necessity modals, all modals *Multi-functional:* BE verbs, predicative adjectives, present tense verbs
Scientific style	Short passives, all passives, definite articles
Coherence and development of arguments	Adverbial conjuncts, conditional subordinating conjunctions, other subordinating conjunctions
Writing prompt	Place nouns

The key features of the MICUSP sub corpus when compared against the SPR corpus. The features (shown in the right column) were grouped thematically and given a label (shown the left column).

Table 8.4 presents four thematic groups: expression of stance, scientific style, coherence and development of arguments, and writing prompt. We describe each theme and provide illustrative text samples.

Expression of Stance

Expression of stance includes the largest number of features. Stance refers to the ways that speakers and writers express their feelings, attitudes, judgments, and assessments in discourse (Biber et al., 1999, p. 966), often through adverbials (e.g., *unfortunately*), modal verbs (e.g., *may*), and complement clauses (e.g., *the fact that…*). While more frequent in spoken language, stance remains an important interpersonal function for academic writers as well (Hyland, 2005).

As Figure 8.1 illustrates, modal verbs were among the most significant of the key features in the MICUSP student papers. Specifically, the necessity modals *should, must,* and *have to*, as well as the prediction modals *will, would, be going to,* and *used to*, were frequent, occurring (in aggregate) 8.3 times per one thousand

words (ptw) in the student writing but just 3.09 times ptw in the science press releases. In the student papers, the functions of these modal verbs were closely tied to the specifics of their writing assignments. To illustrate this, consider the following two excerpts from different student papers. Prediction modals are bolded and necessity modals underlined.

Text Sample #4: MICUSP Student Writing

Wolverine Consultants had put together a construction management plan that **will** guide the team throughout the design and construction of the Wight-Green Building. We **will** be responsible for scope and change order management, keeping track of all change orders and forwarding all information on to subcontractors. An analysis of the bid market **will** be conducted and a packaging strategy **will** be put together. We **will** monitor milestones in the schedule to properly manage the design, bidding and construction schedules.

Text Sample #5: MICUSP Student Writing

Plug-in hybrid electric vehicles (PHEVs) <u>should</u> also be seen as a potential first step towards a long-term pathway to transition to vehicles powered by hydrogen fuel cells. The key drivers for PHEV policies <u>should</u> be: Promoting the adoption of PHEVs is an important component of reacting to all of these drivers but additional policy measures are necessary in order to maximize the economic security and benefit to the US. If U.S. taxpayers are willing to adopt policies that promote PHEVs, there <u>should</u> be some consideration of this adoption as something that will benefit the U.S. economy in the long term.

In Text Sample #4, the purpose of the report is to describe a future project. This demands a greater use of *will* to indicate future time and make predictions. In Text Sample #5, the purpose of the student paper is to present electric vehicles as a potential solution to an energy problem. As a result, the student writer frequently adopts the modal verb *should* to make recommendations for how their implementation could be made. Thus, the characteristics of these writing tasks in part motivate the use of certain modal verbs, reflecting the importance of situational characteristics of language production (Biber & Conrad, 2019).

Another set of key features which played a role in expressing stance in student writing included certain predicative adjectives and the verbs that connect those adjectives to their subjects. Like modal verbs, adjectives can express stance meanings like likelihood, attitude, and evaluation (Biber, 2006). In the student papers, such adjectives were frequently found in the predicative position along with the present tense form of be. To illustrate this, Table 8.5 lists the top ten predicative adjectives in the MICUSP subcorpus. Adjectives expressing stance meanings are bolded.

Table 8.5. Top Ten Most Frequent Predicative Adjectives in the MICUSP Subcorpus*

Predicative Adjective	Normalized Frequency per 1,000 Words
important	0.35
likely	0.20
difficult	0.18
necessary	0.15
possible	0.13
present	0.10
available	0.10
able	0.09
responsible	0.07
similar	0.07

The ten most frequent predicative adjectives in the MICUSP subcorpus in order of frequency (normalized per 1,000 words). Adjectives expressing stance appear in bold.

Six of these predicative adjectives express stance meanings related to likelihood (*likely, possible*), evaluation (*important, difficult*), necessity (*necessary*), or ability (*able*). In student papers, these adjectives were linked to their subjects by way of a copular verb, usually the BE verb, meaning that the three key features of BE verbs, predicative adjectives, and present tense verbs co-occurred to express stance. To illustrate these features in discourse, consider the following excerpt from a student paper. Predicative adjectives expressing stance are bolded, while other stance expressions are underlined.

> **Text Sample #6:** MICUSP Student Writing
>
> Even if the US were to foot the bill for the entire malarial control program in Sub-Saharan Africa ($824 million annually), malaria control <u>would</u> cost 1/6 of the amount the United States spends annually on funding for cancer-research [9]. While cancer research is **important** and <u>will</u> save many lives in the future, malaria control <u>can</u> save millions of lives right now. It <u>can</u> also help bring impoverished countries out of debt and help them become participants in a strong global economy, and provide trade and investment opportunities for American businesses.

Text Sample #6 illustrates how the present tense form of BE can co-occur with predicative adjectives to express stance. The student writer uses the evaluative adjective *important* to characterize the noun phrase *cancer research* by linking the two with the present tense copular verb *is*. Around this expression can also be

found other stance expressions, such as the modal verbs *will* and *can*. Expressions like these were key features of the student writing relative to the science press releases, suggesting that writers of press releases either withhold their stance to a greater extent or express stance in other ways.

▪ Scientific style

Scientific style was characterized by short passives, all passives, and the definite article. The frequent use of passive voice in academic writing has been well researched, in part due to the social functions that the passive fulfills. Specifically, the passive voice is said to diminish the role of the individual researcher on the results in order to portray experiments as replicable by others in their academic community (Ding, 2002). Given the tendency of SPR writers to use narrative style in reporting academic research, it is unsurprising that passive voice verbs were key features of the students' academic writing. Table 8.6 presents the normalized rates of long passives, short passives, and postnominal passives, where long passives include a *by* prepositional phrase indicating the agent of the verb, short passives exclude this phrase, and postnominal passives are nonfinite passives postmodifying nouns.

Table 8.6. Normalized Rates of Three Kinds of Passive-Voice Verbs in the MICUSP and SPR Corpora*

Type of Passive Voice	MICUSP Subcorpus (ptw)	SPR Corpus (ptw)
Long passive	1.99	1.95
Short passive	14.1	8.91
Postnominal passive	3.43	3.87
Total	19.5	14.7

A comparison of the rates of three different kinds of passive–voice verbs in both the MICUSP subcorpus and the SPR corpus. The frequences for each type of passive are normalized per 1,000 words.

Table 8.6 demonstrates that the greater use of passives overall in the student writing is largely due to short passives, as rates of long and postnominal passives were similar between the corpora. Not only did the student writers use more passives, they also used passive verbs that were qualitatively different from those used by SPR writers. For example, while the student writers relied most on BE *used*, the SPR writers relied most on BE *published*. Examples are shown below (1–2):

1. Purple loosestrife plants **are used** to treat health problems such as diarrhea, dysentery, ulcers and sores. [Student paper #106]
2. The study, "Field and full-scale laboratory testing of prototype wildland fire shelters," **was published** online in the International Journal of Wildland Fire. [SPR 1534]

Clearly, the passive form of *publish*, such as in (2), performs a function specific to registers needing to specify the periodical in which an academic study was published. The passive use of *use*, on the other hand, such as in (1), is particularly common in empirical research writing (Millar, Budgell, & Fuller, 2013). In short, SPRs not only use fewer passives overall but also use them for different purposes.

As a final note on the theme of scientific style, we also categorized definite articles (*the*) here. Definite articles are used with either countable or uncountable nouns and signal that the referent of the noun is known to the reader. Due to its nominal style, definite articles are used most often in academic writing (Biber et al., 1999, p. 267), resulting in an informationally dense style of discourse that can convey an academic *feel* to the discourse. The following excerpt from a student paper illustrates this style. Noun phrases determined by *the* are underlined, their head nouns are italicized, and their definite articles are bolded.

> **Text Sample #7:** MICUSP Student Writing
>
> Epilepsy, one of **the** most common neurological *disorders*, is characterized by **the** sudden *onset* of recurrent *seizures* due to **the** synchronous *ring* of *populations* of *neurons*. Due to **the** debilitating *nature* of *seizures* and **the** *fact* that approximately 1% of **the** *population* suffers from *epilepsy*, much *research* has investigated **the** *dynamics* of **the** *onset* of *seizures* with **the** *hopes* of developing *methods* of *seizure prediction* [1]. This *research* relies primarily on **the** *analysis* of *electroencephalogram (EEG) recordings* which record **the** neural *activity* of epileptic *patients*.

This extract from a student paper includes many nouns determined by *the*. While *the* is frequent in part because there are many nouns overall, another motivating factor is the many phrases and clauses postmodifying these head nouns. That is, the nouns can be determined by *the* because their referent is in the material postmodifying them. For example, the head noun of the second bolded instance of *the* above is *onset*. While *onset* has not yet been introduced in the paper and thus should not be determined by *the*, the postmodifying information makes it clear what *onset* the writer is referring to, namely *the onset of recurrent seizures due to the synchronous ring of populations of neurons*. As this example suggests, writing that uses a lot of definite nouns postmodified by phrases and clauses conveys a technical style of discourse, and this style was more often found in the MICUSP student papers than the SPRs.

Coherence and Development of Arguments

The third major thematic group of key features in student-produced science writing is the coherence and development of arguments, which includes linking adverbials and several dependent clauses. Linking adverbials (labeled *conj_advl* in Figure 8.1) refer to phrases that signal the logical relationship between clauses or

phrases. While there are several semantic groups of linking adverbials (see Biber et al., 1999, pp. 875–879), the most frequent linking adverbials in both corpora function to compare and contrast ideas (*however, instead*), present logical results or inferences (*thus, therefore*), and enumerate on ideas (*first, also*), the latter of which was particularly frequent in the student papers. To see these phrases in action, consider the following excerpt (linking adverbials bolded).

> **Text Sample #8:** MICUSP Student Writing
>
> **Furthermore,** the observed mechanisms used by the host species to counteract the parasitism, may force genetic changes that will later make possible a genetically based form of isolation. **However,** they caution that the balance between low error rates and finding a conspecific mate create many difficulties; when colonization does occur the cost of finding a mate that knows the same song as you may be too high to make the colonization viable. **Therefore,** they assume that many colonization events probably occurred in the indigobird species before being successful. Conclusively showing that the observed diversification in indigos is due only to a learned song was beyond the scope of their paper; **however,** this theoretical work only supports the genetic evidence.

Three of the four sentences in Text Sample #8 begin with a linking adverbial, including those enumerating on ideas (*furthermore*), offering logical conclusions (*therefore*), and contrasting ideas (*however*). Such sentence-initial adverbials allow the student writer to develop clear arguments by explicitly marking the relationship between ideas. The relative lack of such connectors in the SPRs suggests that those texts either do not explicitly label the relationships between ideas as often or do not elaborate arguments as frequently.

Like these linking phrases, two kinds of clauses used to organize ideas were also found as key features in the student writing, namely adverbial clauses expressing conditions and their outcomes and those expressing a variety of other relationships, such as time, place, and manner (e.g., *after, while, until, since*). Interestingly, such clausal adverbials have been shown to be characteristic of more formal contexts of spoken English, such as teacher talk in university classrooms (Biber, 2006), and both linking and clausal adverbials are characteristic of lower levels of university-level student writing (Staples et al., 2016). Thus, in contrast to the discussion of passive verbs and definite articles earlier, the key features related to developing arguments found in the MICUSP subcorpus are comparatively representative of student writing.

Writing Prompt

Finally, nouns referring to locations (*nn_place* in Figure 8.1) were key features of student writing. Review of these nouns suggested that they largely relate to

the specific topics of student papers, especially those relating to environmental issues. Nouns such as *forest, environment, field, area*, and *space* were common due to the topics that students chose to write about. Thus, a student corpus with a wider variety of writing prompts may not replicate place nouns as a key feature of that register.

▮ Conclusion

In this study, we set out to identify the key features of scientific press releases relative to student-written scientific reports. Our reason for doing so was to better understand how these two registers differed linguistically. The key feature analysis presented in this study resulted in 22 lexicogrammatical features, ranging from verbal features to nominal features to clausal features. Nine of these features were descriptive of the science press releases, meaning that they were key features due to their being consistently more frequent in SPRs. These nine features of the SPR corpus were arranged into three thematic groups pertaining to information density (longer words, greater lexical diversity), informational reports of past events (past tense verbs, public verbs, communication verbs, proper nouns, common nouns, all nouns), and timeliness (adjectives of time). These features reflect the fact that press releases, despite being written for broader nontechnical audiences, are still informationally dense, highly literate texts, as evidenced by their longer, more diverse words and density of various kinds of nouns. They also reflect one of the main functions of press releases: to report scientific research as news. Writers frequently introduce new actors and their professional affiliations into the report, resulting in many proper nouns. Past tense public verbs are also frequently adopted to report their words.

The remaining thirteen key features were descriptive of student-produced science writing. These features were organized into four thematic groups: expression of stance (modals, BE verbs, predicative adjectives, and present tense verbs), scientific style (passives, definite articles), coherence and development of arguments (adverbial conjuncts, subordinating conjunctions), and writing prompt (place nouns). These features suggest that student-written science reports share features characteristic of scholarly academic writing, namely the use of passive voice, modal verbs to express stance (e.g., attitudes of certainty or uncertainty), and features like adverbial conjuncts and subordinating conjunctions to develop coherence throughout a text. These texts were also constrained by the writing prompt, which is characteristic of student-produced writing.

The results of our study can be used to inform pedagogical guidelines for people who want to communicate science to a popular audience by writing science press releases. This group may include students (both from science and technical communication backgrounds), professional scientists, and practicing technical communicators who frequently work with scientists to communicate expert information to a wider audience. To this end, we conclude this chapter

with takeaways for teachers and practitioners wanting to emulate the language of popularized science press releases. Many of these recommendations follow established guidelines for writing to general audiences. The recommendations we make here are based on the findings presented above.

- **Tell stories using active voice.** Our study found that press releases often included active voice verbs with human subjects as their agents. These linguistic choices are not only stylistic but functional as well. The active voice allows science writers to tell stories with human actors who perform actions, think thoughts, and express language. In contrast, the student reports included more passive voice verbs, especially those ellipsing the agent. Thus, writers who want to conform to expectations of press releases should consider using the active voice when possible and understand how it connects to the purposes or demands of the situation. Joseph Williams (2009, Ch. 2) includes helpful questions that students can use to help determine when and how to use the passive voice clearly.

- **Include others' words by using quotations and communication verbs.** Like journalistic writing, press releases often include direct quotations from experts and other third parties. Incorporating their perspectives into press releases is essential. Linguistic devices for doing this include communication verbs like *say*, *remark*, and *report*, as well as the reporting clause syntax (i.e., *"...", said the researcher.*). These features can help students and practitioners new to press releases integrate others' words and ideas into their texts.

- **Consider avoiding technical jargon.** The science press releases examined in this study relied heavily on nouns, a characteristic feature of literate, informationally dense discourse (Biber & Gray, 2016). This finding was also supported by the higher type-token ratio and word length in the press releases. At the same time, the frequent nouns used in the press releases were decidedly nontechnical, suggesting that students and practitioners should be careful when including technical jargon in their press release writing. Technical terms can be expanded into longer phrases or clauses that explain the meaning intended by the original language.

- **Highlight the importance of research with adjectives of time.** In an environment inundated with more and more research (Hyland & Jiang, 2019), popularizers need to persuade readers that their press releases report on meaningful studies worth reading about. This study suggests that one way science press releases accomplish this is by using adjectives of time like *new*, *increasing*, and *latest* to communicate these findings, which were ubiquitous in our corpus. These offer a simple and compact means to motivate the necessity of one's reporting.

- **Minimize expressions of authorial stance.** While virtually all registers express stance in some way, we found that science press releases use

significantly fewer of two kinds of stance expressions relative to the student papers: modal verbs and predicative adjectives. This can be interpreted as press release writers abiding by the journalistic norms of objectivity and balance (Dunwoody, 2014), which in turn reduces the role of the writer in interpreting and evaluating information. Therefore, we recommend that those learning to write press releases carefully consider the ways they insert their opinions and judgments when communicating research to broader audiences.

- **Report facts rather than make arguments.** The notion that science press releases may make lesser use of stance expressions due to their function to report facts and circumstances rather than make arguments was also born out in the comparatively fewer number of linking adverbials. That is, because the student reports served, in part, to make arguments and flesh out those arguments, they used more linking devices like *first, also*, and *however*. Moreover, compared to (student or expert) research writing, press releases are much briefer, so avoiding unnecessary wording is helpful for meeting length requirements.

The six guidelines presented above could be adopted into an assignment appropriate for a typical introductory technical communication class. An instructor could ask students to find a published scientific report on a topic they are interested in and then write a 500-word press release announcing the study. In addition to describing the rhetorical makeup of press releases, the instructor could also introduce the linguistic features associated with the recommendations in the list above in examples of science press releases and share them with students. Doing this could help students develop their abilities to recognize the linguistic features that distinguish press-release writing from the kinds of science writing many of them have been taught to produce in the past. The features from the list above could also be converted into a rubric used to assess the assignment so students are encouraged to practice using as many of the recommendations as they can in their press releases. An assignment like this one, specifically focused on science communication, could also help students more clearly see science communication as a topic relevant to technical communication (see Carradini, 2020).

Limitations and Future Research

While we believe the findings and resulting recommendations shared above are valuable, they are not without limitation. For example, unlike other studies that compared the content of research articles and their accompanying press releases (e.g., Sumner et al., 2016), we studied a broad sample of press releases and a broad—but unrelated—sample of student-produced science writing. In doing so, we are not able to make claims about how specific content is changed as it is

reconstructed from one context to another. Similarly, since the corpus of student writing lacks granular detail (e.g., on specific writing assignments, classroom settings, etc.), we were limited in our understanding of the situational characteristics for this corpus. Moreover, differences in the formal characteristics of the corpora (i.e., number and length of texts) likely shape our results as well.

However, a comparison like the one presented here is still useful because it identifies specific linguistic features of press releases that differ from features common to the academic writing many students are taught to produce. Teachers can explicitly point to these differences in teaching students to write science popularizations and ask students to use them in their assignments.

Another limitation results from the study's purely textual focus. Because we examined only the texts of student reports and science press releases, we are unable to confidently determine authorial intent or the reasons authors may have chosen to use the linguistic features we observed. Future studies might interview writers of science press releases or students writing popular texts for the first time to better understand how they produce their texts.

In a similar vein, we are also careful to note that we did not assess effectiveness in this study. That is, we identified the linguistic features with frequencies that varied significantly between science press releases and student-produced scientific writing, but we make no claims as to whether using these features will improve the effectiveness of press releases or science popularizations in general. However, the findings presented here provide an initial step in understanding how the language of science writing and science popularizations differ. Future research can build on these findings by conducting experiments (see Graham, 2017) to determine the extent to which the use of these features are effective in communicating to a popular audience.

■ References

Balzotti, Jon, & Hansen, Derek. (2019). Playable case studies: A new educational genre for technical writing instruction. *Technical Communication Quarterly*, *28*(4), 407–421. https://doi.org/10.1080/10572252.2019.1613562

Bawarshi, Anis S., & Reiff, Mary Jo. (2010). *Genre: An introduction to history, theory, research and pedagogy*. Parlor Press and The WAC Clearinghouse. https://wacclearinghouse.org/books/referenceguides/bawarshi-reiff/

Bazerman, Charles. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. University of Wisconsin Press. Available in open access at https://wacclearinghouse.org/books/landmarks/bazerman-shaping/.

Biber, Douglas. (1988). *Variation across speech and writing* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511621024

Biber, Douglas. (2006). University language. In *Scl.23*. John Benjamins. https://benjamins.com/catalog/scl.23

Biber, Douglas, & Conrad, Susan. (2019). *Register, genre, and style* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511814358

Biber, Douglas, Davies, Mark, Jones, James K., & Tracy-Ventura, Nicole. (2006). Spoken and written register variation in Spanish: A multi-dimensional analysis. *Corpora*, *1*(1), 1–37. https://doi.org/10.3366/cor.2006.1.1.1

Biber, Douglas, & Egbert, Jesse (2018). *Register variation online*. Cambridge University Press.

Biber, Douglas, & Gray, Bethany. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press. https://doi.org/10.1017/CBO9780511920776

Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan, & Finnegan, Edward. (1999). *Longman grammar of spoken and written English*. Longman.

Brezina, Vaclav (2018). *Statistics in corpus linguistics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316410899

Burke, Amy, Okrent, Abigail, & Hale, Katherine. (2022, January 18). *The state of U.S. science and engineering 2022*. The National Science Foundation. https://ncses.nsf.gov/pubs/nsb20221

Cagle, Lauren E., & Herndl, Carl. (2019). Shades of denialism: Discovering possibilities for a more nuanced deliberation about climate change in online discussion forums. *Communication Design Quarterly*, *7*(1), 22–39. https://doi.org/10.1145/3331558.3331561

Calsamiglia, Helena, & Van Dijk, Teun A. (2004). Popularization discourse and knowledge about the genome. *Discourse & Society*, *15*(4), 369–389. https://doi.org/10.1177/0957926504043705

Carradini, Stephen. (2020). A Comparison of Research Topics Associated with Technical Communication, Business Communication, and Professional Communication, 1963–2017. *IEEE Transactions on Professional Communication*, *63*(2), 118–138. https://doi.org/10.1109/TPC.2020.2988757

Catenaccio, Paola. (2008). Press releases as a hybrid genre: Addressing the informative/promotional conundrum. *Pragmatics*, *18*(1), 9–31. https://doi.org/10.1075/prag.18.1.02cat

Chavanachid, Pititul, & Sripicharn, Passapong. (2019). A Corpus-informed study of move structures and linguistic features of press release and corresponding news reports of products and services in the Thai business context. *Journal of Studies in the English Language*, *14*(1), Article 1.

Cohen, Joseph. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Earlbaum Associates.

Crone, Wendy C., Dunwoody, Sharon L., Rediske, Raelyn K., Ackerman, Steven A., Zenner Petersen, Greta M. & Yaros, Ronald A. (2011). Informal science education: A practicum for graduate students. *Innovative Higher Education*, *36*(5), 291–304. https://doi.org/10.1007/s10755-011-9176-x

De Oliveira, Janaina Minelli, & Pagano, Adriana Silvina. (2006). The research article and the science popularization article: A probabilistic functional grammar perspective on direct discourse representation. *Discourse Studies*, *8*(5), 627–646. https://doi.org/10.1177/1461445606064833

Ding, Daniel D. (2002). The passive voice and social values in science. *Journal of Technical Writing and Communication*, *32*(2), 137–154. https://doi.org/10.2190/EFMR-BJF3-CE41-84KK

Dunwoody, Sharon L. (2014). Science journalism: Prospects in the digital age. In M. Bucchi & B. Trench (Eds.), *Routledge handbook of public communication of science and technology* (2nd ed., pp. 27–39). Routledge.

Egbert, Jesse, & Biber, Douglas. (2020). It's just words, folks. It's just words. Donald Trump's distinctive linguistic style. In U. Schneider, & M. Eitelmann (Eds.), *Linguistic inquiries into Donald Trump's language: From 'fake news' to 'tremendous success'* (pp.17–40). Bloomsbury Academic. https://doi.org/10.5040/9781350115545

Fahnestock, Jeanne (1986). Accommodating science: The rhetorical life of scientific facts. *Written Communication*, *3*(3), 275–296. https://doi.org/10.1177/0741088386003003001

Gotti, Maurizio. (2014). Reformulation and recontextualization in popularization discourse. *Iberica, 27*, 15–34.

Graham, S. Scott. (2017). Data and lore in technical communication research. *Communication Design Quarterly*, *5*(1), 8–25. https://doi.org/10.1145/3090152.3090154

Granado, António. (2011). Slaves to journals, serfs to the web: The use of the internet in newsgathering among European science journalists. *Journalism*, *12*(7), 794–813. https://doi.org/10.1177/1464884911412702

Halliday, M. A. K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power*. Falmer Press.

Hilgartner, Stephen. (1990). The dominant view of popularization: Conceptual problems, political uses. *Social Studies of Science*, *20*(3), 519–539. https://doi.org/10.1177/030631290020003006

Hyland, Ken. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, *7*(2), 173–192. https://doi.org/10.1177/1461445605050365

Hyland, Ken. (2010). Constructing proximity: Relating to readers in popular and professional science. *Journal of English for Academic Purposes*, *9*(2), 116–127. https://doi.org/10.1016/j.jeap.2010.02.003

Hyland, Ken, & Jiang, Feng (Kevin). (2019). *Academic discourse and global publishing: Disciplinary persuasion in changing times.* Routledge.

Hyland, Ken, & Salager-Meyer, Françoise. (2008). Scientific writing. In B. Cronin (Ed.), *Annual review of information science and technology* (vol. 42, pp. 297–338). Information Today.

Jacobs, Geert. (1999). *Preformulating the news: An analysis of the metapragmatics of press releases.* Benjamins.

Jacobs, Geert. (2014). Press releases. In K. P. Scheider & A. Barron (Eds.), *Pragmatics of discourse* (vol. 3, pp. 583–599). De Gruyter.

Lin, Lifeng, & Aloe, Ariel M. (2021). Evaluation of various estimators for standardized mean difference in meta-analysis. *Statistics in medicine*, *40*(2), 403–426. https://doi.org/10.1002/sim.8781

Liu, Shuai, & Zhang, Jingyuan. (2021). Using metadiscourse to enhance persuasiveness in corporate press releases: A corpus-based study. *SAGE Open*, *11*(3), 21582440211032. https://doi.org/10.1177/21582440211032165

Luzón, María-José, & Pérez-Llantada, Carmen. (Eds.). (2019). *Science communication on the internet: Old genres meet new genres* (Vol. 308). John Benjamins. https://doi.org/10.1075/pbns.308

Maat, Henk Pander. (2007). How promotional language in press releases is dealt with by journalists: Genre mixing or genre conflict? *Journal of Business Communication, 44*(1), 59–85. https://doi.org/10.1177/0021943606295780

Mauzy, Ann. (1994). Shifting our focus: Communicating science to a new, nontechnical culture. *Scaling New Heights in Technical Communication IPCC 94 Proceedings*, 142–146. https://doi.org/10.1109/IPCC.1994.347531

Millar, Neil, Budgell, Brian, & Fuller, Keith. (2013). "Use the active voice whenever possible": The impact of style guidelines in medical journals. *Applied Linguistics*, *34*(4), 393–414. https://doi.org/10.1093/applin/ams059

Mogull, Scott A. (2011). A call for new courses to train scientists as effective communicators in contemporary government and business settings. In J. Conklin, & G. F. Hayhoe (Eds.), *Qualitative research in technical communication* (pp. 189–211). Routledge.

Mogull, Scott A. (2018). *Scientific and medical communication: A guide for effective practice*. Routledge.

Myers, Greg. (2003). Discourse studies of scientific popularization: Questioning the boundaries. *Discourse Studies*, *5*(2), 265–279.

National Academies of Sciences, Engineering, and Medicine. (2017). *Communicating science effectively: A research agenda*. National Academies Press. https://doi.org/10.17226/23674

Navarro, Danielle (2015). *Learning statistics with R: A tutorial for psychology students and other beginners (Version 0.6)*. University of New South Wales.

Nordquist, Richard. (2020, March 26). *What is lexicogrammar?* ThoughtCo. https://www.thoughtco.com/what-is-a-lexicogrammar-1691120

Parkinson, Jean, & Adendorff, Ralph. (2004). The use of popular science articles in teaching scientific literacy. *English for Specific Purposes*, *23*(4), 379–396. https://doi.org/10.1016/j.esp.2003.11.005

Paul, Danette. (2004). Spreading Chaos: The role of popularizations in the diffusion of scientific ideas. *Written Communication, 21*(1), 32–68. https://doi.org/10.1177/0741088303261035

Pelger, Susanne. (2018). Popular science writing bringing new perspectives into science students' theses. *International Journal of Science Education, Part B*, *8*(1), 1–13. https://doi.org/10.1080/21548455.2017.1371355

Pelger, Susanne, & Nilsson, Pernilla (2016). Popular science writing to support students' learning of science and scientific literacy. *Research in science education*, *46*(3), 439–456. https://doi.org/10.1007/s11165-015-9465-y

Peters, Hans Peter, Brossard, Dominique, de Cheveigné, Suzanne, Dunwoody, Sharon L., Kallfass, Monika, Miller, Steve, & Tsuchida, Shoji. (2008). Interactions with the mass media. *Science*, *321*(5886), 204–205. https://doi.org/10.1126/science.1157780

R Core Team. (2020). *R: A language and environment for statistical computing*. https://www.r-project.org/

Samraj, Betty. (2016). Research articles. In K. Hyland, & P. Shaw (Eds.), *Routledge handbook of English for academic purposes* (pp. 403–415). Routledge.

Schwartz, Lisa M., Woloshin, Steven, Andrews, Alice, & Stukel, Therese A. (2012). Influence of medical journal press releases on the quality of associated newspaper coverage: Retrospective cohort study. *BMJ*, *344*, d8164–d8164. https://doi.org/10.1136/bmj.d8164

Schwitzer, Gary (2008). How do US journalists cover treatments, tests, products, and procedures? An evaluation of 500 stories. *PLoS Medicine*, *5*(5), e95. https://doi.org/10.1371/journal.pmed.0050095

Spoel, Philippa, Goforth, David, Cheu, Hoi, & Pearson, David. (2008). Public communication of climate change science: Engaging citizens through apocalyptic narrative explanation. *Technical Communication Quarterly*, *18*(1), 49–81. https://doi.org/10.1080/10572250802437382

Staples, Shelly, Egbert, Jesse, Biber, Douglas, & Gray, Bethany. (2016). Academic writing at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication*, *33*(2), 149–183. https://doi.org/10.1177/0741088316631527

Stryker, Jo Ellen. (2002). Reporting medical information: Effects of press releases and newsworthiness on medical journal articles' visibility in the news media. *Preventive Medicine*, *35*(5), 519–530. https://doi.org/10.1006/pmed.2002.1102

Sumner, Petroc, Vivian-Griffiths, Solveiga, Boivin, Jacky, Williams, Andrew, Bott, Lewis, Adams, Rachel, Venetis, Christos A., Whelan, Leanne, Hughes, Bethan, & Chambers, Christopher D. (2016). Exaggerations and caveats in press releases and health-related science news. *PLOS ONE*, *11*(12), e0168217. https://doi.org/10.1371/journal.pone.0168217

Swales, John (1981). *Aspects of article introductions*. University of Aston.

Swales, John (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

The Regents of the University of Michigan. (2009). *Michigan Corpus of Upper-level Student Papers* [Data set].

Weber, Ryan (2020). The news from Mars. *Technical Communication Quarterly*, *29*(2), 136–154. https://doi.org/10.1080/10572252.2019.1689298

Wells, Rulon (1960). Nominal and verbal style. In T. A. Sebeok (Ed.), *Style in Language*, (pp. 213–220). MIT Press.

Wickham, Hadley (2016). *ggplot2*. Springer. https://doi.org/10.1007/978-3-319-24277-4

Williams, Joseph M. (2009). *Style: Lessons in clarity and grace*. (9th Ed.). Longman.